DOCUMENT RESUME

ED 386 975 HE 028 556

AUTHOR Patrick, William J.; Stanley, Elizabeth C.
TITLE Assessment of Research Quality. AIR 1995 Annual Forum Paper.
PUB DATE May 95
NOTE 24p.; Paper presented at the Annual Forum of the Association for Institutional Research (35th, Boston, MA, May 28-31, 1995).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; Departments; Doctoral Programs; Educational Assessment; *Evaluation Criteria; Foreign Countries; Government School Relationship; Higher Education; *Institutional Research; Program Effectiveness; *Quality Control; *Research; Researchers; *Research Universities
IDENTIFIERS *AIR Forum; Performance Indicators; *United Kingdom; University of Glasgow (Scotland)

ABSTRACT

The British experience of nationwide research quality assessments and new measures to improve accountability are described. Consequences for the higher education system and for individual institutions are examined. Three national assessments of the quality of research in the United Kingdom (U.K.) in 1986, 1989, and 1992 have provided a standard rating of research programs in order to increase selectivity in allocating funds. In each assessment institutions prepared information for assessment by central panels, which judged the usefulness of certain statistics as indicators of research quality. Comprehensive peer review of research quality has demonstrated the usefulness of certain performance indicators. Information which is collected about faculty researchers is identified, including number of publications produced and number of research assistants. Issues related to quality ratings of departments are addressed. Efforts of the University of Glasgow (Scotland) to maintain and enhance its research quality ratings are discussed. Comparisons are made of U.K. research assessment approaches and U.S. research doctorate program assessment using measures of the National Academy of Sciences. Implications of U.K. approaches to assessment for U.S. institutional researchers are addressed. Generally, the British results give support to the use of some quantitative measures of research outputs and inputs as indicators of overall research quality; however the system would be unlikely to transfer successfully into a new setting if the underlying purpose diverges in any way from the British experience. (Contains 16 references.) (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 386 975

Assessment of Research Quality

William J. Patrick
Planning Officer
University of Glasgow
University Avenue
Glasgow G12 8QQ
Scotland
44-141-330-4218
Fax: 44-141-330-4920
E-mail: bill@mis.gla.ac.uk

Elizabeth C. Stanley
Director of Institutional Research
Iowa State University
215 Beardshear Hall
Ames, IA 50011-2030
515-294-1181
Fax: 515-294-4835
E-mail: estanley@iastate.edu

Paper to be presented at the 35th Annual Forum of
The Association for Institutional Research
Boston, Massachusetts
May 28-31, 1995

IR 020 556

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Fifth Annual Forum of the Association for Institutional Research held at the Boston Sheraton Hotel & Towers, Boston, Massacusetts, May 28-31, 1995. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Jean Endo
Editor
AIR Forum Publications

Abstract

The assessment of research quality and accountability for research funding are major issues in higher education. This presentation describes the British experience of nation-wide research quality assessment exercises, and newly-introduced measures intended to improve accountability. The consequences are examined, including those for the higher education system as a whole and for individual institutions. Comprehensive peer review of research quality has demonstrated the usefulness of certain performance indicators. Their relevance for United States institutional researchers will be discussed, and similarities with the measures used by the National Academy of Sciences in its assessment of research-doctorate programs will be identified.

Introduction

Assessment, accountability and the evaluation of program effectiveness are increasingly important to institutions of higher education throughout the world and to those who support them. This paper considers the British experience of nation-wide research quality assessment exercises and their consequences for both individual institutions and the higher education system as a whole.

Research effort in the United Kingdom forms only a small part of the world total. For example, in 1991, the U.K.'s share of world science papers was 8.6% (Institute for Scientific Information, 1992). The U.K. must use its scarce resources to best effect, if it is to compete internationally. The selective use of public funds to support high-quality research has been a recurrent theme in Government thinking over many years. Accountability for the use of research funds is a further concern of Government. An explicit commitment to wealth creation and the quality of life by more effective use of basic research is another recurrent theme, dating back to the beginning of the 19th century, and recently restated (Parliamentary Office for Science and Technology, 1993). The requirements for increased accountability and, more particularly, four-yearly national assessments of research quality are sometimes considered to be symptoms of increasing Government intrusiveness in university affairs, and they have had a profound impact on academic life in the U.K. in recent years.

Accountability for the Use of Research Funding

Since the mid-1980s, it has been the policy of the central higher education funding bodies in the U.K. to direct financial support to areas of research excellence, identified in regular national assessments of research quality. While each university has autonomy concerning the use of its central government appropriations, there is nonetheless an expectation that resources allocated on research-based criteria should be used primarily for research, rather than teaching, which is financed separately. Tension between the desire to provide some direction concerning the use of research funding, while at the same time not requiring research appropriations to be spent in each institution exactly as allocated by the central funding bodies, has led to an increased emphasis on the need for individual institutions to account for the ways in which they use research funds.

Various methods for such accounting have been considered (Coopers & Lybrand, 1993). A process was sought by Government which would both achieve better accountability for the use of public funds and assist institutional management by increasing awareness of the cost of research. It was also considered desirable to achieve better public (and parliamentary) understanding of what universities actually do.

It was recognized that a system in which institutions would be required annually to account for the use of research funds could have certain pitfalls. It was feared that the cost, in terms of academic time lost and antipathy aroused,

could be such as to outweigh the benefits. Debate focused on whether accountability should be in terms of actual expenditure by institutions or in terms of the funds which they allocate internally for research. The former solution was recognized as necessitating a diary exercise as a means of determining the extent to which each individual faculty member spent her or his time on research activities. The allocation model was the one which was adopted, because it is less burdensome. Institutions now report once a year on how they intend to allocate research appropriations among subject groupings.

The Assessment of Research Quality

In addition to requiring accountability for the uses of research funds, the Government has conducted three national assessments of the quality of research in the U.K. – in 1986, 1989 and 1992 – to provide a standard rating of research programs in order to increase selectivity in the allocation of central government funding. Each exercise has been more refined than the previous one. Institutions of higher education are now preparing for a fourth research assessment exercise in 1996.

In each exercise, institutions prepare standard statistical and narrative material for submission to central panels for assessment. The information considered by the panels is purely documentary; there are no site visitations, for example, and no appeals against the decisions of the panels are allowed. The provision of information to the assessors is a daunting task for administrators, particularly in the large, research-oriented universities. However, these exercises provide a unique opportunity to assess the usefulness of certain statistics as indicators of the quality of research.

In the 1992 national research assessment exercise, each institution could make a submission in any of 72 units of assessment, defined to cover a full range of academic disciplines, including clinical medicine and dentistry (Universities Funding Council, 1992a). Information provided by the institutions was used as the basis for peer review by specialist panels, which awarded quality ratings on a five-point, criterion-referenced scale (Universities Funding Council, 1992a):

5 Research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others.

4 Research quality that equates to attainable levels of national excellence in virtually all sub-areas of activity, possibly showing some evidence of international excellence, or to international level in some and at least national level in a majority.

3 Research quality that equates to attainable levels of national excellence in a majority of the sub-areas of activity, or to international level in some.

2 Research quality that equates to attainable levels of national excellence in up to half of the sub-areas of activity.

1 Research quality that equates to attainable levels of national excellence in none, or virtually none, of the sub-areas of activity.

Each of the 72 units was assessed by a panel consisting of appropriate specialists. Preeminence in research was the main criterion for appointment to the panels. Members were appointed by the central funding body, the Universities Funding Council, acting on the advice of subject associations, learned societies, professional associations, and relevant industrial and business organizations. Panels operated on a "core plus adjunct" model, whereby the core panel (typically 5 to 8 members) could be joined by a number of specialist advisers to cover specific areas.

The panels considered all types of research: applied, strategic and basic. In 1992, these terms were defined as follows (Universities Funding Council, 1992a):

> Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.
>
> Strategic research is applied research which is in a subject area which has not yet advanced to the stage where eventual applications can be clearly specified.
>
> Applied research is work undertaken in order to acquire new knowledge. It is, however, directed primarily towards practical aims or objectives.

The requirement to distinguish in science and engineering between basic/strategic research, on the one hand, and applied research, on the other, arose from the Funding Council's desire to ensure that full recognition should be given to work of direct relevance to the needs of commerce, industry and the public and voluntary sectors.

The information provided included demographic data for each faculty member included in the assessment, the total number of publications produced by each individual over the last three and a half years (four and half years in arts and humanities), and the number of research assistants and postgraduate research students supervised. Additionally, each faculty member cited her or his two best publications over the same period, as well as up to two forms of other public output.

Numerical summaries of research publications for each department were divided into 12 categories:

1. Authored Books
2. Edited Books
3. Short Works (a category which was criticized as being too wide, and which included chapters in books, letters to journals, notes in journals, short papers, abstracts, official reports, review articles, monographs, pamphlets, entries in encyclopedias and dictionaries, and departmental working papers)
4. Conference Contributions, refereed (excluding abstracts, and various contributions to conference proceedings not subsequently published)
5. Conference Contributions, other (subject to the same exclusions as refereed conference contributions)
6. Editorships (in spite of doubts as to whether they constitute research)
7. Papers in Academic Journals
8. Papers in Professional Journals (The distinction drawn between academic and professional journals was fuzzy. Academic journals were considered to contain research papers aimed primarily at the academic and research community. The papers would normally, but not always, be refereed. A professional journal was considered to be aimed at practicing members of a profession, containing papers informing members of current developments in their profession rather than research results. Such papers would not normally be refereed by the academic and research community.)
9. Papers in Popular Journals
10. Reviews of Academic Books
11. Other Publications (including invited papers and, in the creative and performing arts, original works such as scores and scripts)
12. Other Public Output (including databases, architectural and engineering designs, maps, patents granted, copyrights, license agreements, publicly-available software and, in the creative arts, work performed or created and publicly recognized as original contributions to art, sculpture, drama or music)

Additional data for each department included the numbers of postgraduate research students, research studentships and degrees awarded, and the numbers, amounts and sources of its external research grants and contracts.

While much of the information provided was factual, some of it was intended to allow the "personality" of the unit to be assessed. Departments were invited to describe, in one page, the context in which their research was taking place, and the mechanisms they had to promote, manage and monitor research. Departments were also invited to describe their plans for the future in a single page. A catch-all "general observations" section was included, to allow departments to describe any special factors relevant to the exercise which might otherwise be overlooked. The assessors paid particular attention to these free-text submissions, suggesting that there was a significant element of subjective judgment applied (Higher Education Funding Councils, 1993).

The Results of the 1992 Research Assessment Exercise

In the 1992 research assessment exercise, 172 institutions of higher education were assessed (Universities Funding Council, 1992b), and there was considerable variation among institutions in the ratings. There was also a significant divergence in performance between the country's traditional universities, and the new universities (formerly polytechnics and colleges of higher education), which first became eligible for assessment in 1992. Figure 1 shows two series: one for the traditional universities, and another for the new ones. Each point on the series represents an institution's (unweighted) average percentage score. There is an interesting overlap in the two series. The considerable overall variation in the results is to some extent a natural consequence of diversity in institutional missions.

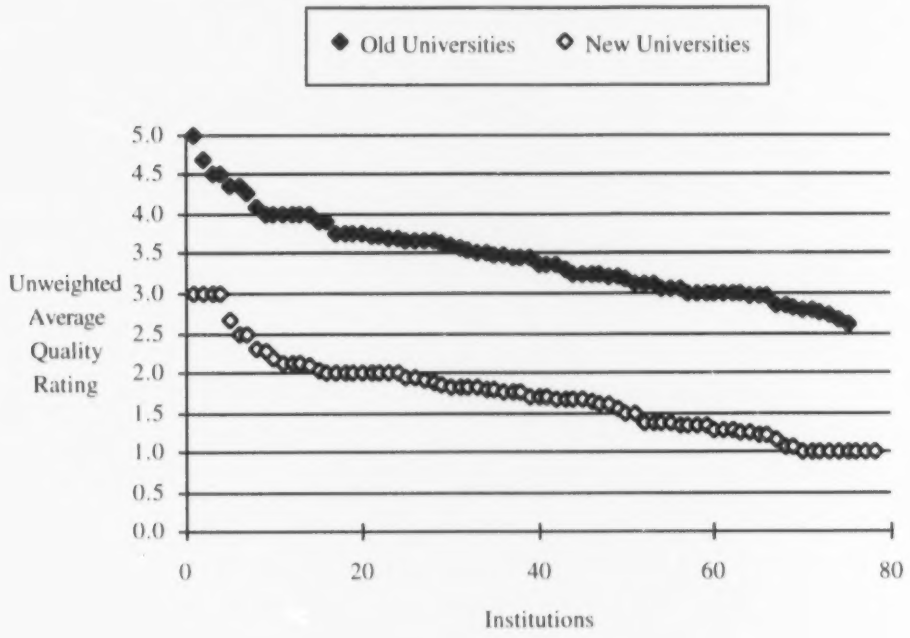

Figure 1 – Variation in Quality Among Institutions

Within institutions, ratings for individual disciplines tended to vary. Figure 2 illustrates, for 12 large urban universities, the percentage of units having each of the five quality ratings. The performance of the best institutions is certainly impressive, but the general picture is of a system in which the large, research-oriented universities have a mixture of highly-rated and lesser-rated departments. Any presumption of uniform quality within or among these institutions would be fallacious.

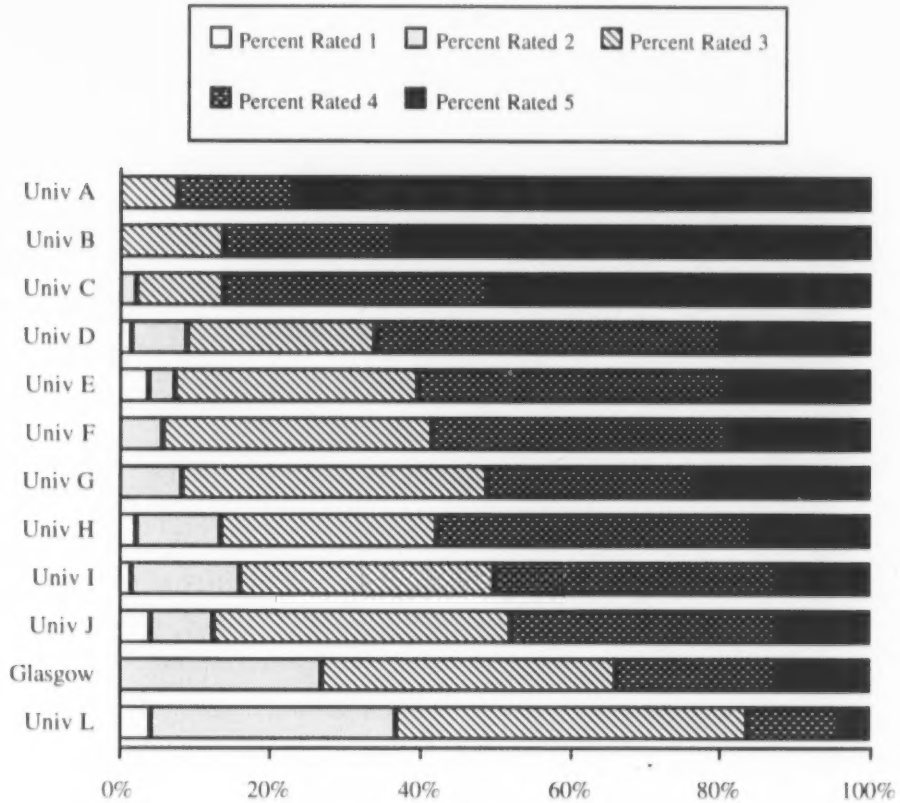

Figure 2 – Variation in Quality within Large, Research-Oriented Universities

Among disciplines, average ratings varied substantially. To a significant extent, this was due to the proportion of new universities, which are generally more oriented toward teaching, in each unit of assessment (Taylor, 1995). The ratings were criterion-referenced, with no normalization process used. It is not impossible to suppose that some assessment panels raised the mean research rating in their subjects, knowing that this could increase the allocation of funds to them, albeit at the expense of others. It has been suggested (Taylor, 1995) that competition between panels could lead to an overall inflation of ratings, and it seems apparent that differences among mean ratings should be considered in any interpretation of the results. Figure 3 shows the frequency distribution of the average scores of the 72 units of assessment. Anthropology, genetics and classics were rated as British universities' top three research areas in 1992, with average scores of 4.29, 4.17 and 4.15 respectively. The lowest rating was for nursing, which had an average score of 2.11.

BEST COPY AVAILABLE

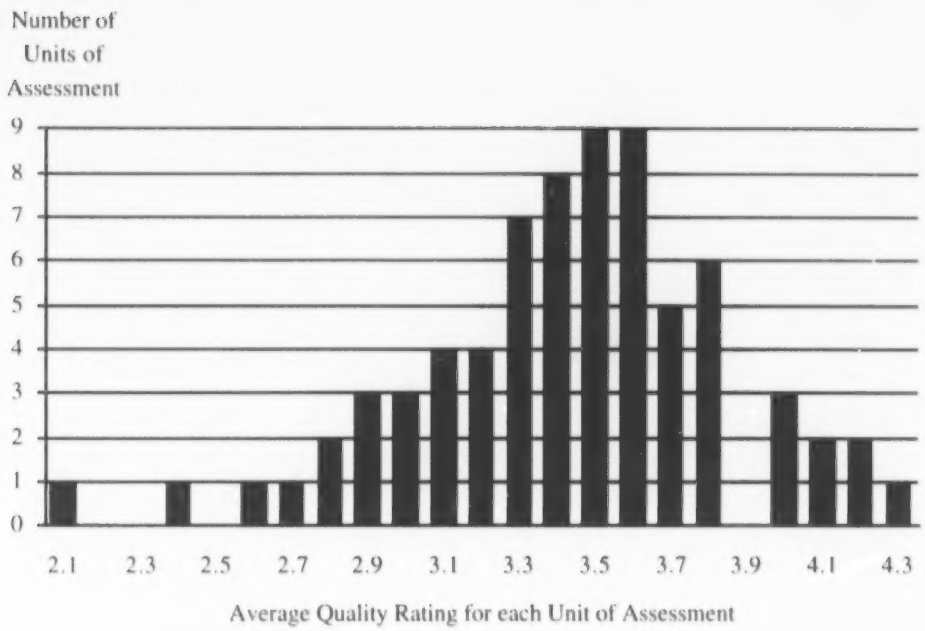

Figure 3 – Frequency Distribution of Average Quality Ratings

Within disciplines, the ranking of departments is all-important. Research reputations stand or fall on how one compares with peer departments at other institutions. For example, in nursing a 3-rating would have gained a respectable place in the top twenty percent of the country's research departments. On the other hand, the same rating in genetics would have placed a department at the bottom of the rankings.

The Consequences for the Higher Education System

The most immediate and significant effect of the national research assessment exercises has been in the level of baseline funding for research allocated to each of the higher education institutions by the central funding bodies. The degree of concentration of resources in areas of research excellence is such that virtuous and vicious cycles are likely to become apparent over time (Cradden, 1995). Those institutions which are better off are able to invest more in centers of research excellence; those with less money find it increasingly difficult to maintain their standing. The stratification between "research" universities and others is likely to become more pronounced over time. Competition for research "stars" is intensifying and there is likely to be an exodus of talent from poorly-rated departments towards their more illustrious counterparts. The award of external peer-reviewed research grants earmarked for specific purposes is likely to become more concentrated in those institutions having better research facilities and infrastructure. Graduate student entrants are likely to give preference to highly-rated departments.

BEST COPY AVAILABLE

In the 1989 exercise, the work of all faculty members was taken into account, but in 1992 each institution had to decide for itself whether the work of particular individuals was sufficiently meritorious to justify their inclusion for assessment. There is a trade-off built into the funding mechanism: a larger number of faculty included in the exercise attracts more money; but more prestigious research also implies more money. Within most departments, there is a "tail" of relatively unproductive faculty whose inclusion in the exercise will only bring down the quality rating for the department as a whole. The precise nature of the trade-off is not known until after the results of the exercise are known – a matter of much contention for those in the invidious position of deciding which colleagues are worthy of assessment, and which are not.

Figure 4 shows the trade-off in financial terms between quality and quantity, as currently practiced by the Higher Education Funding Council in England. The trade-off became known only when the Funding Council announced its allocations to institutions for the 1993-94 fiscal year. It adopted the so-called 'J-1' formula. A quality rating of J received a financial weighting of J-1, and 1-rated departments were not funded at all. The percentage of faculty members assessed was summarized by use of a letter, as illustrated on the z-axis of Figure 4. Quality ratings (2 through 5) are shown on the x-axis, and relative financial pay-offs are represented by displacement along the vertical (y) axis. For this figure, the financial weighting for a unit with a research rating of 2A was arbitrarily set at 1. As the figure demonstrates, there were several anomalies in the funding formula. Some of these are indicated by the shading of blocks in the figure. Not all 5-rated departments were necessarily better funded than 4-rated departments, for example. Prior knowledge of the system might have allowed an institution to achieve a better financial result by receiving a 4A rating rather than a 5C or a 5D rating.

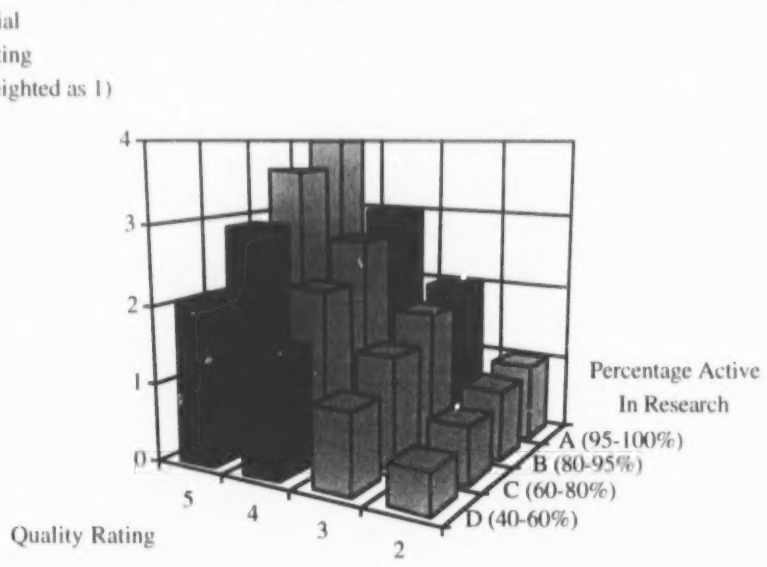

Figure 4 – The Trade-Off between Quality and Quantity

A focus on only the details of the funding formula is undoubtedly too simplistic, however. The unquantifiable factor of public approbation is virtually unaffected by the percentage of faculty assessed. In practice, research reputations depend on the numerical quality ratings. On these rest future promotion prospects for individuals, and their ability to attract external research funding. The capacity of departments to attract and retain top-class researchers also depends on the prestige which high quality ratings bring.

The Consequences for Individual Institutions

This section reports on experience within one of the major research universities – the University of Glasgow – based on plans prepared by its constituent academic planning units, to which substantial financial and academic responsibilities are devolved. In 1993, there were 22 such planning units, each made up of groups of cognate academic departments, and constituting the middle level in the University's managerial structure. A comparison with other reports (Williams, 1991, 1993) suggests that Glasgow is typical of the research-oriented institutions in the higher education sector.

The description of Glasgow as a research university simply identifies one of its distinguishing characteristics. In fact, its mission is tripartite: both teaching and outreach are also included. This paper describes only the University's efforts to maintain and enhance its research quality ratings, although the University also strives to be rated highly in the Scottish Higher Education Funding Council's assessment of the quality of teaching and learning. Different institutions place different emphases on research and teaching, and at Glasgow it is contended that there is synergy between the two.

Many of the phenomena described below existed at Glasgow in embryonic form before the research assessment exercises took place, and they might have developed over time, even without the added stimulus provided by increased external scrutiny. Nonetheless, it is the belief of the authors that what is described is to a considerable extent a product of the new, highly competitive environment in which research takes place. The issue for institutional researchers and policy makers is whether the behavior patterns described are conducive to strategic intent, or whether they are irrelevant or even dysfunctional. Undoubtedly, "the climate has become cooler and the planning keener" at Glasgow as new approaches to funding have led to a more competitive atmosphere (Temple & Whitechurch, 1994).

The University of Glasgow's responses to selectivity in research funding may be represented as having three main dimensions: organizational, financial, and human resources.

Institutional Responsiveness – The Organizational Dimension

The organizational dimension of Glasgow's response has five parts: increased monitoring and control of research; greater coordination of research activity; the quest for greater efficiency through reorganization; the encouragement of a research culture; and the promotion of the University's research activities, both internally and outside the University.

Glasgow has had a central research committee since the mid-1980s. Lower-level committees have been established more recently, both at the planning unit and departmental levels. There are now three tiers of committees responsible for promoting, monitoring and evaluating research quality and output, as well as defining future research directions.

Planning unit research committees are responsible for monitoring departmental research strategies and levels of activity. In particular, the progress of research grant applications and success rates in obtaining external research funding are monitored. At least three planning units have engaged external consultants to provide reports on their present standing and their future prospects in research. The major goal is to obtain quality ratings in the 1996 research assessment exercise which are at least as high as in the previous exercise. This entails giving advice on strategy and tactics for the 1996 exercise.

The intention is to maximize collaboration and the exploitation of expertise and facilities both within planning units, and across the University as a whole. While the University's overarching research committee takes responsibility for coordinating University-wide research efforts, planning unit research committees ensure that departmental strategies are consistent with planning units' own research plans. They encourage interdisciplinary research initiatives, for example, through the establishment of centers of research excellence, and the promotion of seminars as a means of allowing members of staff and research students to exchange views on research work and techniques, and the sharing of databases. They also coordinate research efforts within their particular disciplines, and encourage interaction with others, both within the University and outside. The national research selectivity exercises have focused on the quality of research at the group level. It is therefore not surprising to find an emphasis within the University on organizing research using group structures, and the encouragement of teamwork in specifically-identified areas. Many planning units have emphasized the need for a more proactive approach in order to initiate and extend campus-wide collaborative work.

The quest for greater efficiency through reorganization has several aspects. In some cases, departments have been merged, and in others the breadth of research coverage has been significantly reduced. More positively, weak areas have been helped either by strengthening them or, if this has not been possible, by suggesting ways whereby they can join with stronger, related groups, where a more meaningful contribution can be made. There has been a concentration of resources and effort in those areas identified as being of high potential. This has been achieved in

some areas by the creation of larger planning units, permitting economies in the levels of administrative support. Research efforts are generally being concentrated on relatively small numbers of research themes which encompass existing and proposed cognate research strengths, focusing intellectual resources in defined areas of front-line research.

There has been widespread encouragement of a high-quality research culture. It is expected that all departments and a large majority of faculty members within them must devote a significant part of their professional efforts to research. A positive attitude toward research among all faculty is seen as the key to increasing postgraduate research student numbers, research income and publication rates, as well as influencing external perceptions of the quality of the University's research and the all-important research ratings themselves. Steps have been taken to encourage a feeling of commitment and unity of purpose in the large research groups which are now emerging, and, equally importantly, to identify clear targets and objectives. Periodic meetings take place at various levels which allow all faculty to contribute to the development of strategy. Regular reporting and monitoring of progress to peer groups and seminars ensure that expertise in current developments is properly disseminated.

Efforts are also being made to promote the visibility and reputation of the University's research efforts. This is done through regional seminar programs, national summer schools, electronic networks and the encouragement of visiting researchers to collaborate with Glasgow workers in cognate fields. Additionally, prospectuses have been developed to attract more research students, and advertise a range of facilities and research support to a wider audience.

Institutional Responsiveness – The Financial Dimension

The financial dimension of Glasgow's response has four parts: a marked increase in internal selectivity in the distribution of research funding, the provision of startup funding for new initiatives; a concerted effort to increase research funding and to diversify its sources; and the reallocation of resources.

All planning units stress the need to give high priority to the support of successful research departments and successful groups within departments. It is also recognized that selective support for individuals with strong research profiles is needed, and that additional budgetary resources are required for high-caliber researchers. Selectivity can take various forms: extra funds may be allocated for equipment, consumables, teaching assistance, and technical assistance. Several planning units emphasize the importance of providing additional funds for visitors and for travel in order to foster wider collaboration. A few planning units separate funding for teaching and research in their allocation procedures, in order to clarify lines of accountability. Some planning units emphasize the importance of pump-priming funds for new initiatives, and the encouragement of younger scholars in establishing their research reputations.

It is recognized that it is important to improve the flow of information to departments concerning opportunities for external research funding. Several central offices have recently been established within the University to facilitate this process, focusing on the maximization of funding from the European Union, industrial and commercial overhead recoveries, and maximum financial assistance from the Research Councils in defraying the cost of the University's research infrastructure, including technical support. Regular reporting and monitoring of progress to peer groups and seminars facilitate the sharing of expertise in the preparation of research project proposals. Efforts are being made to coordinate applications for major items of new equipment, and to identify new initiatives in research applicable to planning units as a whole. In-house expertise in writing grant applications is being pooled by some planning units, both by ensuring that applications are written in consultation with senior members of the relevant research group, and by exploiting the knowledge of faculty who are members of external grant-giving bodies.

Reallocation of resources is common, particularly in terms of space utilization and the allocation of technical support to research groups rather than to individuals. The common purchase, use and maintenance of equipment are further aspects of the drive to make more effective use of resources.

Institutional Responsiveness – Human Resources

The human resources dimension of Glasgow's response to increased external scrutiny of research has three main aspects: the need to make the very best new appointments; the encouragement of older faculty members who are not productive in research to retire early; and the exploitation of various means by which better use can be made of existing faculty time.

Priority is given to investment in human capital. There is a careful review of all candidates for academic appointments to ensure that faculty with the highest possible academic and research potential are recruited. At the other extreme, older individuals who do not participate fully in research are encouraged to seek early retirement. Others may be required to assume additional teaching and administrative duties. Certain faculty no longer receive research funding from their planning units. In some cases, it is intended that technical support will be withdrawn from faculty no longer considered active in research, and allocated to others.

Various methods are being used to make better use of existing faculty time. Much of this is done through the redistribution of workloads. Generally, the key interface between teaching and research will be maintained, with all faculty contributing to both teaching and research, albeit in variable proportions. Considerable effort has been invested in reviewing the workload of individuals, in order to ensure an equitable division of the overall workload. The main opportunity for improving research performance is by freeing up time for research on the part of those who have a proven record of achievement in research, and by allowing rising research talents to flourish. As a matter of policy, teaching loads are distributed non-uniformly. Many planning units invest substantially to provide teaching

assistants to help more senior faculty with their tutoring and marking loads. It is planned that the number of research students should increase significantly.

Planning units have been active in promoting study leave, and paying temporary assistants to cover at least part of the resultant shortfall in teaching effort. The advantages of study leave are very important in terms of learning new techniques, changing research direction, and introducing new teaching ideas.

Considerable effort has been expended in assessing the research performance of individuals. At least one planning unit requires clear statements regarding research work in progress, grants requested and obtained, and future strategy from every faculty member. Individuals are also subject to review by external specialists in their field. Encouragement is given to improving research publications, and faculty are asked to consider carefully the caliber of journals to be used for articles, and to concentrate only on those with a high international peer-review reputation. In one planning unit, every faculty member is expected to be associated with at least one research grant application every year, either singly, or as part of a research team.

Correlation of Quality Ratings with Data Used for the Assessment of Research Quality

Interest has naturally focused on the correlation between the quality ratings awarded and the raw data supplied for the purposes of the 1992 research assessment exercise. The former, having been established through a process of rigorous peer review, should stand as a benchmark of quality against which the various quantitative measures of research may be tested.

A joint performance indicators working group (Higher Education Funding Councils' Joint Performance Indicators Working Group [JPIWG], 1994) has recently concluded that:

1. The total number of publications per researcher was significantly related to research ratings in a majority of units of assessment.

2. The number of academic journal articles per researcher was significantly related to research ratings across a wide range of subjects.

3. With only a few exceptions, the numbers of books, edited books and short works per researcher were significantly related to research ratings only in social sciences, the arts and humanities, and in performing and creative arts (including design and media and communication studies).

4. The number of reviews of academic books per researcher was significantly related to research ratings primarily in non-science subjects (with a few exceptions).

5. The number of refereed conference papers per researcher was significantly related to research ratings in nearly all engineering subjects. It was also significantly related to research ratings in several non-science subjects.

6. The numbers of doctoral students, FTE postgraduate research students, Research Council/British Academy studentships and total research studentships per researcher were significantly related to research ratings across the entire subject range.

7. Total research income and Research Council/British Academy grants per researcher were both significantly related to research ratings across a wide range of subjects, especially in science-based subjects and in the social sciences but less so in the arts and humanities, and in the performing and creative arts.

On the other hand, five of the twelve types of research publications for which data were provided were found to be significantly related to research ratings in only a handful of cases:

Other conference contributions
Papers in professional journals
Papers in popular journals
Other publications
Other public output

A correlation analysis was supplemented by a stepwise regression procedure intended to illustrate the extent to which variation in the quality ratings within each unit of assessment could be explained by variation in the data submitted. On average across all units of assessment, 76% of the variation in research ratings was explained by five or fewer explanatory variables. The percentage of variation explained was over 80% in half of the units of assessment (Higher Education Funding Councils' JPIWG, 1994). The absolute size of the unit assessed appears to have been the single most influential regressor. This bears out the contention held over a number of years that, in research, big is beautiful. (For example, the Merrison Report (Advisory Board for the Research Councils, 1982) recommended that universities should seek to form associations between departments within individual universities, between universities, and with other researchers, and should take the initiative in areas where there was not already a viable research base.) Other variables with statistically significant coefficients for a number of assessment units were articles in academic journals, total external research income, postgraduate research students, short works and books (all divided by the number of researchers). Different variables were more significant in some subjects than others.

It has been proposed that the observed correlations should form the rationale for the central collection, on an annual basis, of research performance indicators for all institutions of higher education in the U.K. Criticism of this proposal has focused on the balance between the usefulness and the cost of obtaining such information. Given that there are to be full-blown, nation-wide research assessment exercises once every four years, it is not clear that it is necessary to obtain annually an enumeration of all research publications. In requiring an annual census, much faith is being placed on correlations which happen to have existed in the past between publications productivity and research quality. The production of particular volume indicators will quickly downgrade their usefulness. The "publish or perish" mentality is already familiar enough in academia. At the same time, the main change in the data requirements for the 1996 Research Assessment Exercise (RAE) is that submissions will not be required to include a summary count of publications. This change has been made because "the funding bodies wish to signal clearly that the RAE is concerned with research quality, and the number of publications and other forms of assessable output is not considered necessarily to be an indicator of research quality" (Higher Education Funding Councils, 1994). Instead, institutions are to be invited to list, for each faculty member, up to four works which will form the basis for the judgment of research quality.

Comparisons with the Assessment of Research-Doctorate Programs in the United States

It is instructive to compare and contrast the U.K. research assessment exercises with the assessment of U.S. research-doctorate programs (Jones, Lindzey and Coggeshall, 1982) sponsored by the Conference Board of Associated Research Councils.

Activities Assessed

The U.K. exercises are designed to assess research, which is defined as "original investigation undertaken in order to gain knowledge and understanding. It includes scholarship; invention and generation of ideas, images, performances and artifacts including design where these lead to new or substantially improved insight; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction" (Universities Funding Council, 1992a). For the purposes of the 1992 research assessment exercise, this definition was further expanded to distinguish basic, strategic and applied research. The purpose of the U.S. exercise was to assess the quality of doctoral programs and, in particular, their effectiveness in preparing students for careers in research. In the U.K., over 2,700 submissions, covering 172 insitutions of higher education, were made in 72 units of assessment, whereas in the U.S., 2,699 graduate programs in 32 disciplines at more than 200 doctorate-granting universities were assessed.

The Definition of Quality

In the U.K., the quality ratings are criterion-based, and are determined by a relatively small number of generally acknowledged experts on the basis of standardized submissions received. The U.S. exercise was conducted by a study committee appointed by the Conference Board of Associated Research Councils and emphasized the importance of a multidimensional approach using a variety of quality-related measures. In the U.K., only a single measure of quality is published (although in some cases, separate ratings were produced in 1992 for applied research as opposed to basic/strategic research). In the U.S., no attempt was made to establish composite rankings of programs either within disciplines or among disciplines. In the U.K., both types of ranking are possible.

The 16 measures reported in the U.S. study included four based on perception ratings (reputational ratings), and it was stated that "impressions of quality, if widely shared, have an imposing reality of their own, and therefore are worth knowing about in their own right" (Jones et al, Social & Behavioral Sciences, 1982, p. 200). However, generally-held perceptions are not a part of the U.K. attempts to assess research quality.

Stakeholders

In the U.K., the principal stakeholders are unquestionably the central government agencies which fund research. Commerce, industry, and the public and voluntary sectors therefore benefit indirectly. Institutional management benefits from being given a clear indication of where research strength lies, both internally and throughout the system. Faculty and students also have an interest in the outcomes of the research assessment exercises. In the U.S., the major stakeholders are seen to be students and their advisors, scholars of the educational system, administrators, policymakers and funders. The U.S. assessments are geared more to the needs of individuals than the U.K. exercises.

Procedure

In the U.S., a study committee identified 16 measures judged relevant to quality in research-doctorate programs. These included the results of a survey of faculty perceptions of the scholarly quality of program faculty, the effectiveness of the program, and recent improvements in program quality. Standardized data are reported by institution and department, and summary statistics are presented for each discipline. In the U.K., relevant information is supplied by the universities to experts, who assimilate it by whatever means they wish, and then announce their conclusions, which are binding on all concerned until the next exercise, using a simple five-point scale.

In the U.S., experts' judgment is used primarily at the planning and design stage of the process and in the submission of reputational survey data. In the U.K., their influence is apparent mainly at the last stage of the exercise, when the quality ratings are decided. The U.K. exercises require active participation on the part of those

being assessed; in the U.S. the necessary information is obtained both from the institutions and from governmental agency or association compilations. The U.K. procedure involves non-academic assessors; the 1982 U.S. exercise did not, although it was felt that this would be advantageous in the future.

The U.S. assessment reported in 1982 had not been repeated until recently. The publication of the results of a new assessment, using similar procedures and measures but expanded to include 41 disciplines, is expected in Summer, 1995. The U.K. exercise was conducted in 1986, 1989, and 1992, and is planned for 1996. The procedures have been refined each time.

Data

After several iterations since 1986, the U.K. system has evolved to the point where the information used for the purposes of assessment of research is reasonably standardized and relatively economical to produce. This information includes both input and output measures as listed in earlier sections of this paper.

In the 1982 U.S. assessment there was an explicit recognition of the limitations of prior studies based primarily on ratings of perceived quality provided by peers – the so-called reputational ratings. One of the purposes of the study was therefore to extend the range of measures used in evaluative studies of graduate programs. These measures include input and output measures representing program size, characteristics of graduates, library size, research support and publication records, as well as the results of a reputational survey.

Relevance for U.S. Institutional Researchers

The British results give support to the use of some quantitative measures of research outputs and inputs as indicators of overall research quality, as determined by the process of peer review. Certain output measures are indicative of quality in their own right, because they are subject to peer review. Papers submitted to academic journals, papers submitted to refereed conferences, applications to and awards from peer-review research sponsors (usually government-financed), and applications for and awards of funding for research studentships are all examples of performance indicators which should have general applicability. It is also of interest to note that much of the variation between institutions in research ratings is accounted for by only five or six explanatory variables. An analysis of the variables found to be significant in the various units of assessment can assist U.S. institutional researchers in the design of institutional processes for the assessment of research activities. However, since there are disciplinary differences among journals, conferences and funding sources, the discriminatory powers of such performance indicators will be greatest when applied across relatively narrow subject areas.

The usefulness of the general U.K. approach to the assessment of research quality is more questionable. The purpose of the British research assessment exercises is quite clear: to achieve selectivity in research funding. The principal stakeholders are also clearly defined: the governmental agencies which fund research. The underlying ethos is to support proven winners at the expense of others. If any of these three facets – purpose, stakeholders, or the underlying system of values – diverge from those pertaining to the British experience, then the system would be unlikely to transfer successfully into a new setting. However, there has been an increasing emphasis on accountability for funds in the United States, and it is probable that more institutions and agencies will attempt to develop their own processes for the assessment of research activities.

The U.S. Government's National Performance Review calls for the use of performance objectives and results in budget and management reviews, and the Government Performance and Results Act of 1993 (GPRA) requires agencies to develop strategic plans, establish performance goals, and report annually on program results, outputs, service levels and outcomes. Agencies are to specify performance indicators or alternative forms of performance goals. The National Science Foundation is conducting a pilot project to develop performance indicators to measure the results of federally funded research programs (Long, 1994). The Federal Office of Science and Technology Policy is engaged in a project on Metrics of Fundamental Science and has sponsored reports and workshops to consider the issues involved in the evaluation of fundamental research programs. Representatives of a number of research-related federal departments and agencies are meeting in a roundtable forum to discuss the development of performance measures for research programs. These activities suggest that U.S. institutional researchers may be increasingly involved in activities related to the assessment of research quality.

The concept of continuous quality improvement is prevalent in the U.S. higher education system, but less so in the U.K. where the emphasis is more on accountability for the use of research funds. Accountability is about value for money, and reduces quality to the lowest common denominator of cash, coupled with the threat of its withdrawal in times of financial stringency. The consequences are likely to include a narrowing of diversity, the reward of conformity, barriers to the development of young researchers, and a stifling of serendipity. The process of improving research quality seems to be a secondary consideration, and the kind of reactive organizational behavior described above is much more likely to prevail.

References

1. Advisory Board for the Research Councils. (1982). Report of a joint working party on the support of university science research. London: Her Majesty's Stationary Office.

2. Coopers & Lybrand. (1993). Research Accountability. Bristol: Higher Education Funding Council for England.

3. Cradden C. (1995). The research "poverty trap". Association of University Teachers Bulletin, April 1995, 4-5.

4. Higher Education Funding Councils. (1993). A report for the Universities Funding Council on the conduct of the 1992 research assessment exercise. Bristol: Author.

5. Higher Education Funding Councils (1994). 1996 Research Asssessment Exercise. Circular Letter RAE96. Bristol: Author.

6. Higher Education Funding Councils' Joint Performance Indicators Working Group (JPIWG). (1994). Consultative Report: Management Statistics and Performance Indicators in Higher Education. Bristol: Author.

7. Institute for Scientific Information. (1992). Science Watch. Vol 3, No 6.

8. Jones, L.V., Lindzey, G. & Coggeshall, P.E. (eds.). (1982). An assessment of research-doctorate programs in the United States. Washington, D.C.: National Academy Press.

9. Long, J. (1994). NSF tries to measure research productivity. Chemical & Engineering News, October 3, 26.

10. Parliamentary Office of Science and Technology (1993). Realising our Potential – A Strategy for Science, Engineering and Technology. London: Her Majesty's Stationary Office.

11. Taylor, J. (1995). A statistical analysis of the 1992 research assessment exercise. Journal of the Royal Statistical Society, Series A, 158, Part 2, 241-261.

12. Temple, P. and Whitechurch, C. (1994). The new entrepreneurship in British higher education. Planning for Higher Education, 22, 13-18.

13. Universities Funding Council. (1992a). Research Assessment Exercise 1992. Circular letter 5/92. Bristol: Author.

14. Universities Funding Council. (1992b). Research Assessment Exercise 1992: The Outcome. Circular letter 26/92. Bristol: Author.

15. Williams, B. (1991). University responses to research selectivity. London: Centre for Higher Education Studies (CHES) of the Institute of Education, University of London. No. 2 of CHES Policy Series.

16. Williams, B. (1993). Research management for selectivity and concentration – with what effect? Higher Education Quarterly, 47(1), 4-16